

Grilles de calcul

Plan

- ? Introduction
- ? Une grille de calcul
- ? Choix infrastructureux
- ? Réalisation pratique
- ? Évolution du middleware
- ? Conclusion

Introduction

Avant la grille...

Equipe ICPS du LSIIT

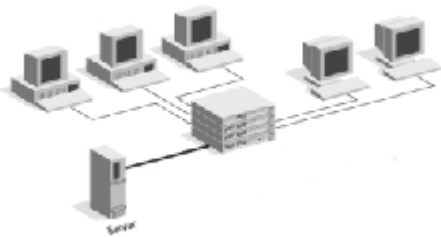
? Recherche en parallélisme

? Administration d'un équipement de calcul/visualisation



Introduction

Mode d'accès aux ressources en réseau



Réseau d'entreprise

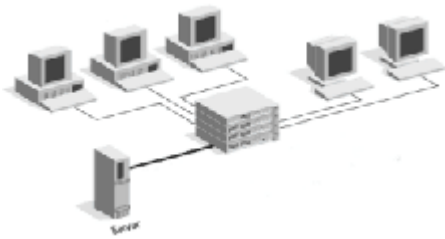
- ? Systèmes de fichiers
- ? Applications
- ? Puissance de calcul



Internet

- ? Fichiers
- ? Messages

Introduction



Utilisateurs :
chercheurs disposant de codes
parallèles

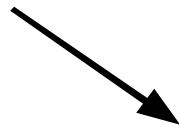
Ressources réparties
hétérogènes

Une grille de calcul

Grille d'ordinateurs

Infrastructure :

- ? Logicielle et matérielle
- ? Efficace, uniforme
- ? Persistante
- ? Peu coûteuse

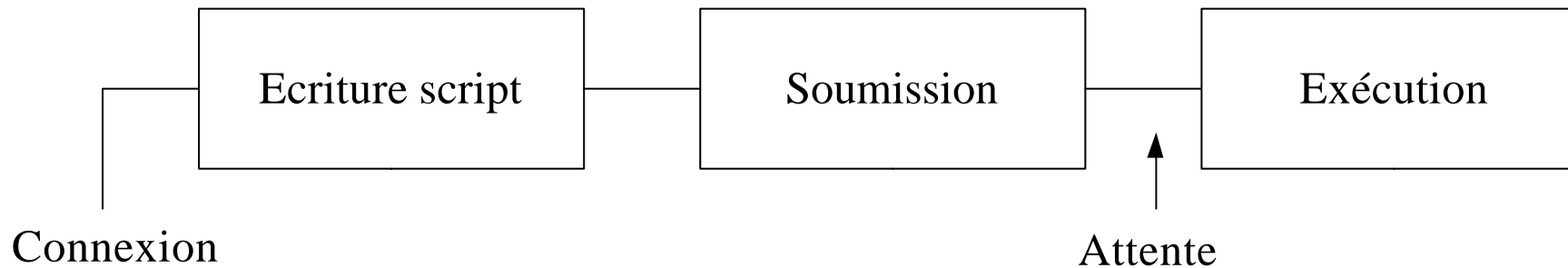


Fournissant l'accès à un ensemble de ressources qui peuvent être éloignées géographiquement.

Une grille de calcul

Traitement par lots (Centres de calcul)

- ? Utilisation de systèmes de files d'attente + réservation de ressources
- ? Machines multiprocesseurs



Accès interactif (Machines de laboratoire)

- ? Utilisation en temps partagé
- ? Stations de travail ou machines multi-processeurs

Une grille de calcul

Quelles contraintes ?

- ? Environnements matériels
- ? Machines existantes
 - ? Réseaux publics
- ? Environnements logiciels
 - ? logiciels libres
 - ? Héritage des codes : non-modification des codes existants
 - ? Non-pénalisation des utilisateurs non-grille
 - ? Sécurité informatique
 - ? Administration autonome de la grille
 - ? Performances ?

Une grille de calcul

Quels critères d'évaluation ?

- ? Fonctionnalité de l'environnement
- ? Commodité d'utilisation : soumission aisée de travaux
- ? Possibilité d'exécution de problèmes réels
- ? Absence de gêne occasionnée aux administrateurs / utilisateurs

Choix infrastructuraux

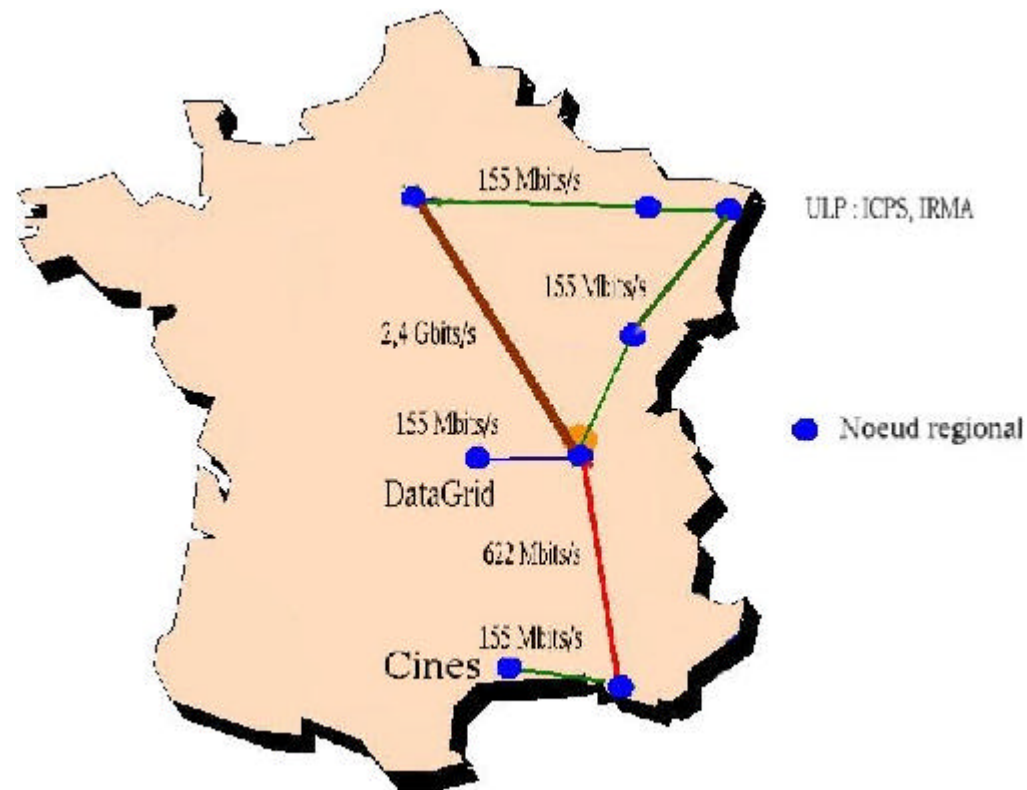
Réseaux

- ? Réseaux d'interconnexion internes
- ? Réseaux locaux
- ? Réseaux métropolitains
- ? Réseaux nationaux

Machines

Nom	Localisation	Processeurs
<i>seven</i>	Illkirch	52×Mips
6 PC divers	"	1×Intel
<i>pellinore</i>	"	2×Intel
<i>irmasrv2</i>	Strasbourg (site Irma)	4×Sparc
<i>irmasrv3</i>	"	12×Sparc
2 PC divers	Strasbourg (site Eost)	1×Intel
PC	Strasbourg (site IECS)	1×Intel
<i>biogrid01</i>	Clermont-Ferrand	4×Intel
<i>minerve</i>	Montpellier	512×Mips

Choix infrastructuraux



Choix infrastructureux

Environnement	Libre	MPI	Portable	Extensible	LSF	Codes	Sécurité
Arrayd	-	✓	-	-	-	✓	-
Mpich	✓	✓	✓	-	-	✓	-
Pacx-MPI	✓	✓	✓	-	-	✓	-
Légion	-	✓	✓	✓	-	✓	✓
Condor-G	-	✓	✓	✓	✓	✓	✓
Corba	✓	-	✓	✓	-	-	-
XtremWeb	✓	-	✓	✓	-	-	✓
Netsolve	✓	-	✓	✓	✓	-	✓
Unicore	-	-	✓	✓	✓	✓	✓
Sun Grid Engine	-	-	-	✓	-	-	✓
Globus	✓	✓	✓	✓	✓	✓	✓

Réalisation pratique

Connexion : identification/authentification

- ? Identification : Nom distinctif (nom de login ne passe pas à l'échelle)
- ? Authentification : clé publique/clé privée

Machines / Utilisateurs identifiés/authentifiés + **certifiés**

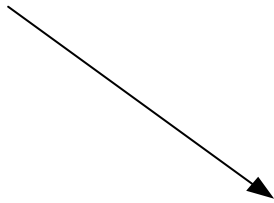
Reconnus par un tiers de confiance

- ⇒ Définition de l'*espace de nommage* des entités
- ⇒ Création d'une autorité de certification locale

Réalisation pratique

Rôle de l'autorité de certification

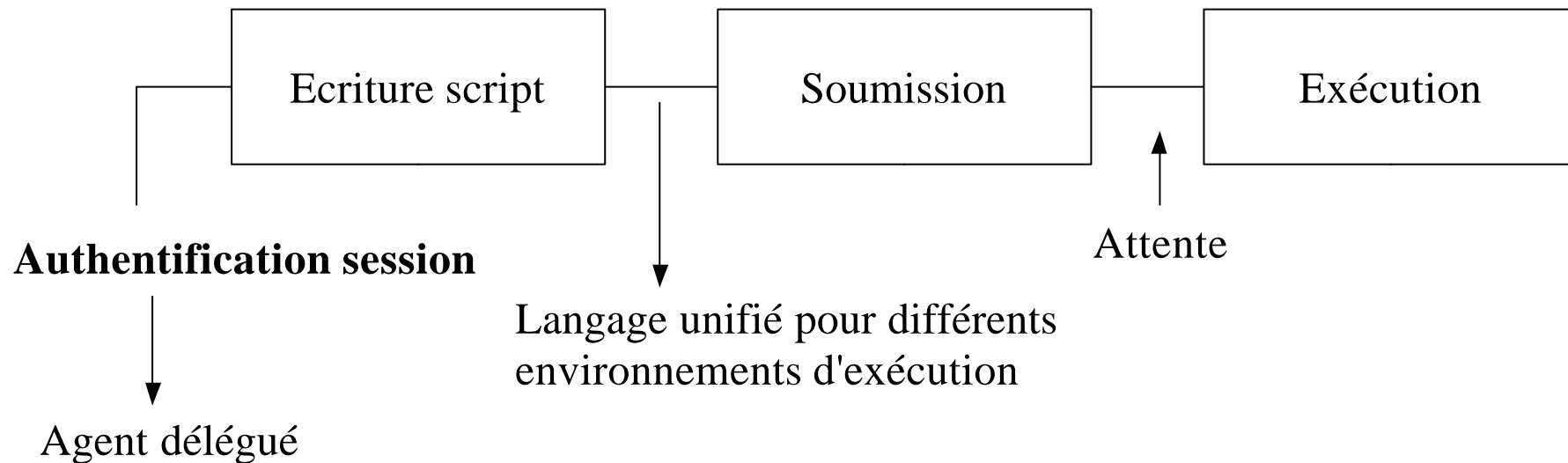
- ? Fournir les certificats aux utilisateurs
- ? Reconnaissance sur la grille : acceptation sur plusieurs machines
- ? Porte d'entrée sur la grille



Permet l'utilisation de la grille : soumission de travaux
+ souplesse lors d'expérimentations

Réalisation pratique

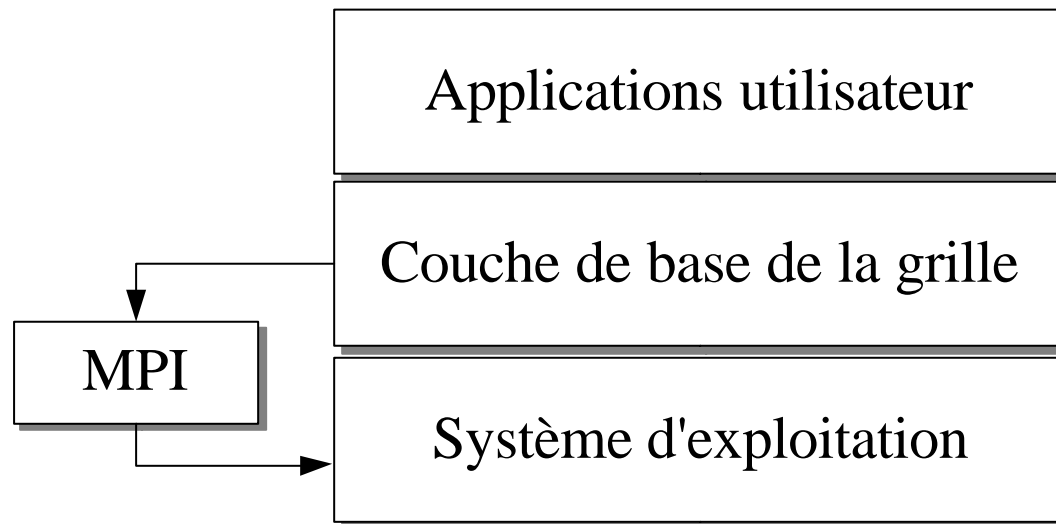
Intégration des systèmes de files d'attente existants



Réalisation pratique

Couche MPI au dessus de Globus

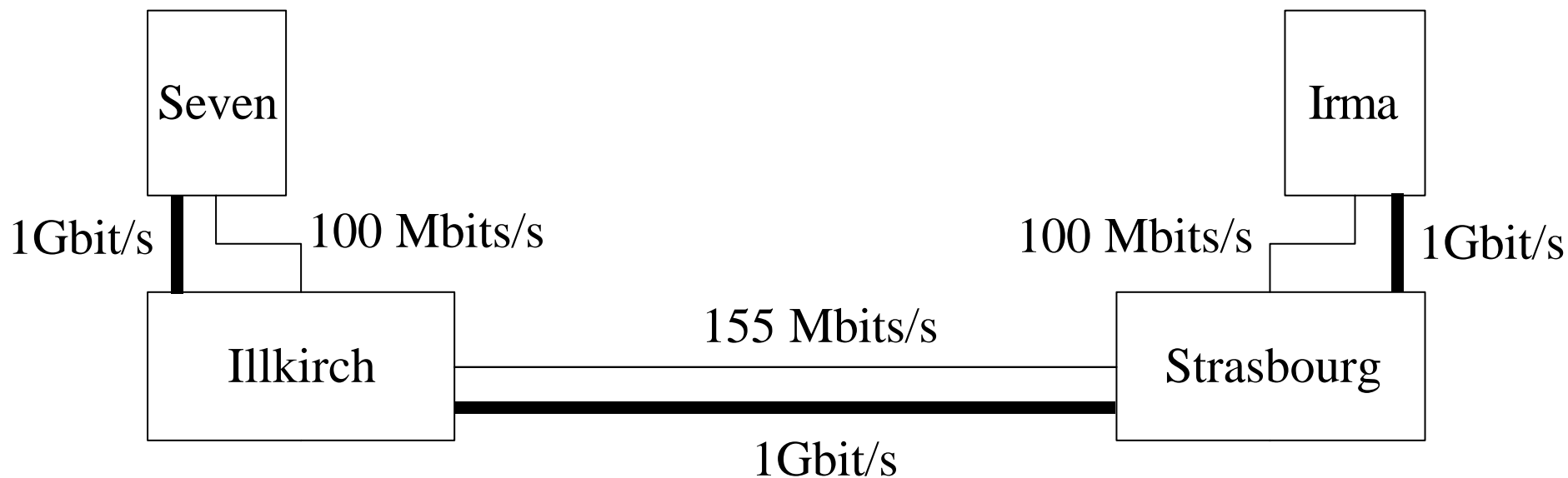
- ? Utiliser les bibliothèques MPI optimisées
- ? Communications inter-machines



Réalisation pratique

Pertes liées à Globus ?

- ? Étude sur réseau métropolitain : Mpich-P4 / Mpich-G2
- ? Jeux d'essais : NPB



Réalisation pratique

Tests de performance

Test	Osiris	Gigabit
FFT / 8	115,52	24,87
Parallèle / 8	15,15	15,17

Temps d'exécution en secondes

Réalisation pratique

Performance relative face à une machine parallèle

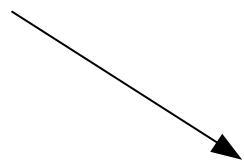
Test	Irma	Seven R12k	Gigabit
FFT / 8	10,70	19,84	24,87
Parallèle / 8	14,82	10,41	15,17

Réalisation pratique

Temps d'attente dans les files

Utilisateur	Nb. travaux		Durée		Temps d'attente		Temps de restitution	
bielecki	198	56	6h36	6h07	1h24	7h00	8h00	13h07
meuwly	108	49	2h55	3h42	2h48	7h34	5h43	11h06

Temps d'attente dans les files pour 8 et 16 processeurs



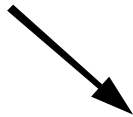
Gain sur temps de restitution

Réalisation pratique

Fonctionnalités spécifiques et recherche

But : Etude d'une méthode d'équilibrage de charge

Moyen : Utilisation des mécanismes de temps partagés



Augmentation du nombre de processus par processeur

Comparaison avec code séquentielisé

Réalisation pratique

Résultats

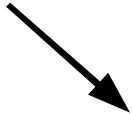
Taille	Original	T. Partagé	Séquentialisation
32	342	154%	88%
48	2005	111%	93%
64	5380	88%	82%

Temps d'exécution en secondes pour 3 tailles

Évolution du middleware

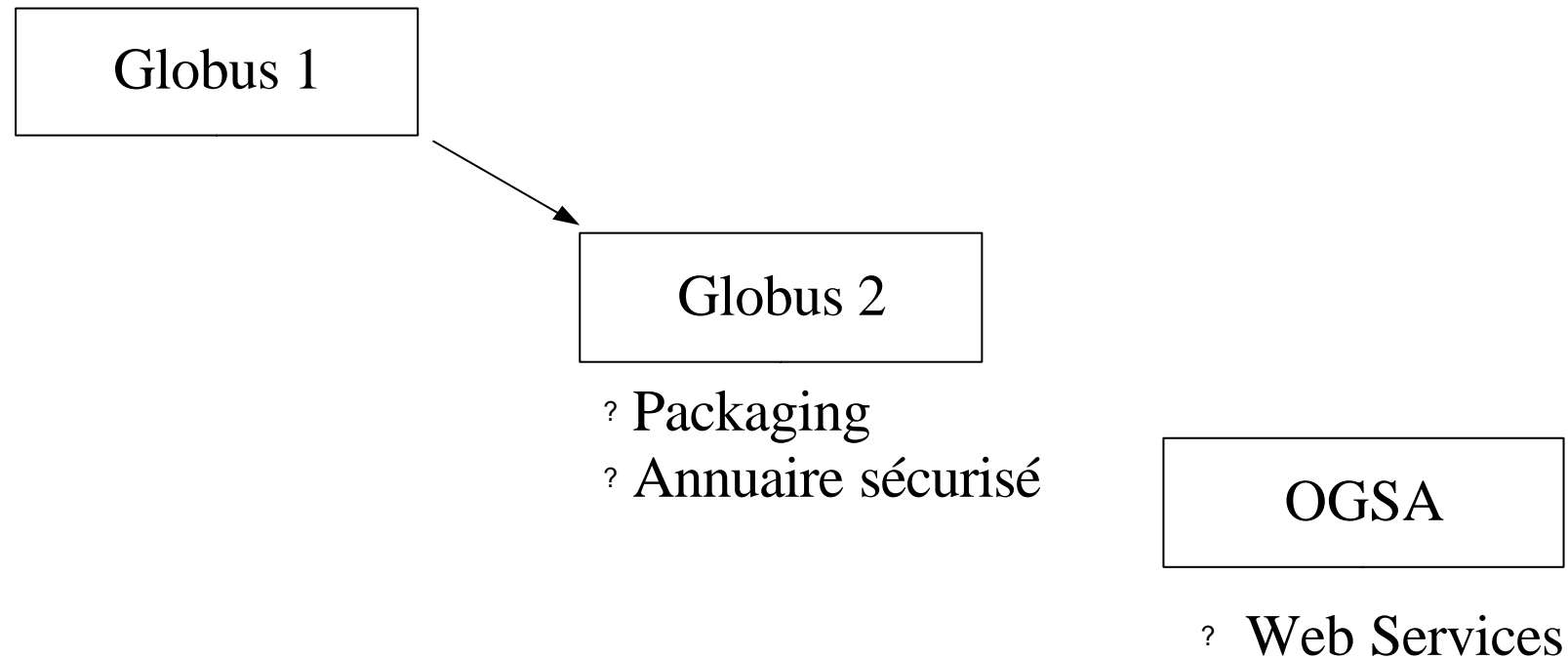
Globus = toolkit

- ? Fournit une API complète pour l'exécution de travaux
- ? Gère bien les aspects d'authentications (Single Sign-On)
- ? Fonctionne !



Insuffisant pour un utilisateur final, équivalent aux lignes de commandes Unix

Évolution du middleware



Évolution du middleware

OGSA : Open Grid Service Architecture
Services Web dédiés à la grille

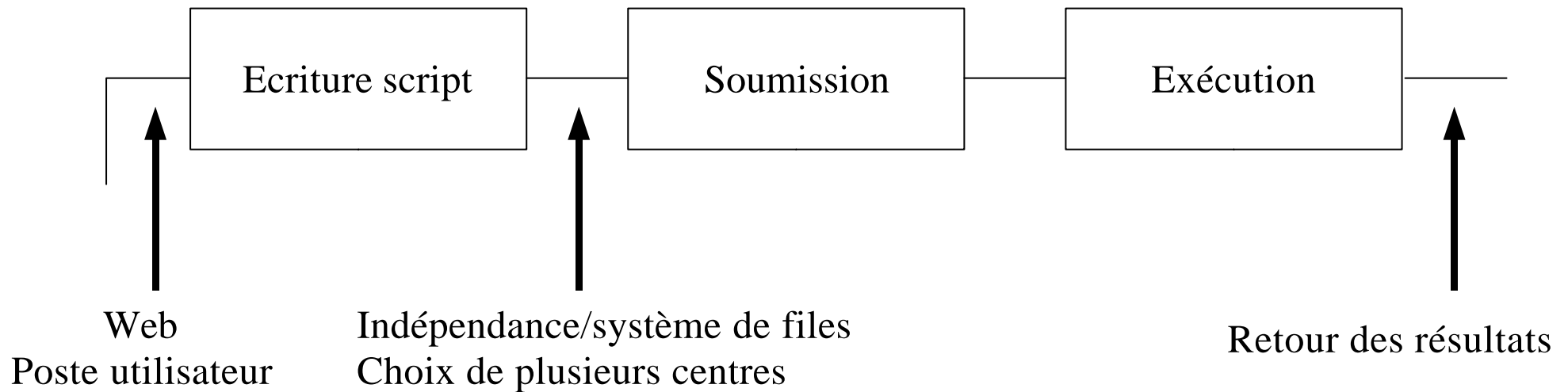
Ensemble de composants OGSI (Open Grid Service
Infrastructure)

But : interopérabilité des grilles

Globus 3 = une implémentation de plusieurs composants OGSI

Conclusion

Évolutions



Conclusion

- ? Grille représentative et efficace
- ? Applications MPI sur machines et réseaux existants
- ? Administration autonome
- ? Utilisation fructueuse en recherche
- ? Choix de l'infrastructure confirmé
- ? Évolutions vers utilisateurs et grille nationale